Amplifying Science Research Through Computational Sciences

Juan Meza

Department Head

Lawrence Berkeley National Laboratory

http://hpcrd.lbl.gov/~meza

AGEP, Rice University **July 16, 2004**

Amplifying the advancement of science and engineering research

Information technology amplifies research in other disciplines in a similar way. As this committee is aware, information technology gave rise to new tools for performing research, computational science techniques. Previously, research was experimental, observational or analytical. Progress in computer and information science and engineering not only advances information technology itself, but leverages advancement of knowledge in other areas. It shares this trait with mathematics. But most other disciplines like astronomy or geology do not offer such leverage.

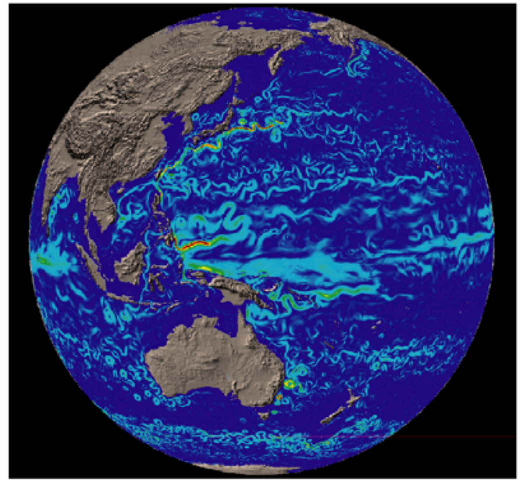
So, my first conclusion is *that investment in the research in computer and information science and engineering has strengthened our economy not just by enabling entirely new products and industries, but by amplifying the efficiency and productivity of almost all other areas of our economy as well as amplifying the advancement of science and engineering research*. It is extraordinarily productive.

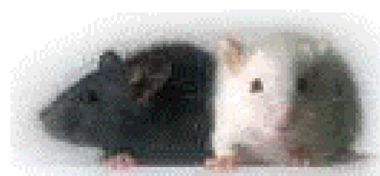
Anita Jones
Quarles Professor of Engineering & Applied Science
University of Virginia
Before the Subcommittee on Research
House Committee on Science
June 16, 2001

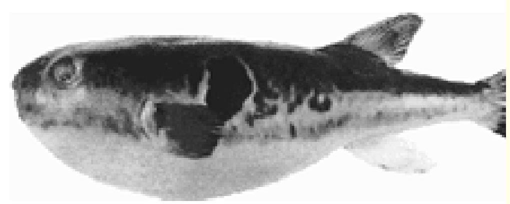
Outline

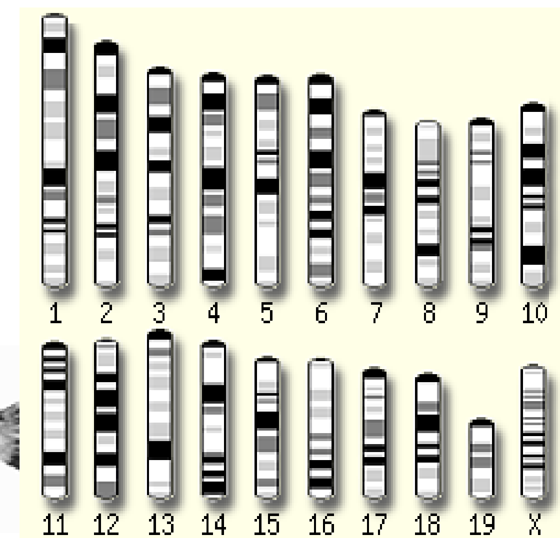
- Short tour of computational science problems
- Role of computational mathematics
- Some optimization challenges
- Summary and lessons learned
- Future directions

Ever broader use of computational sciences for scientific discovery

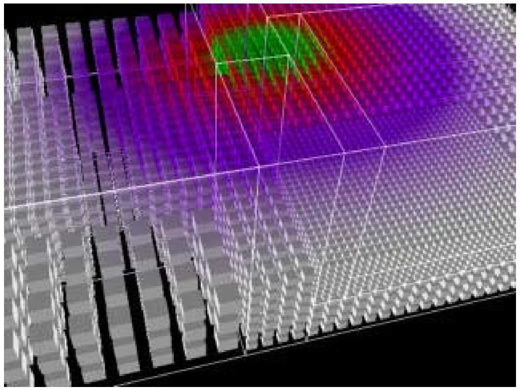


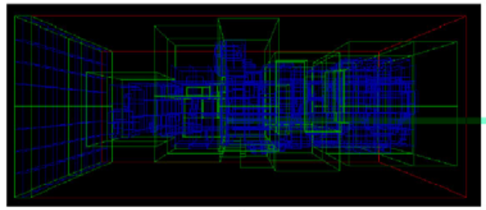




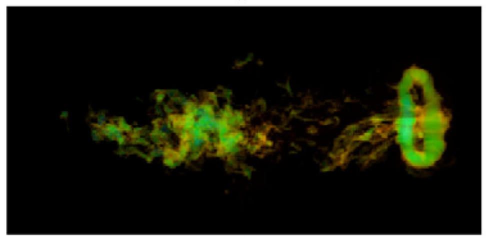


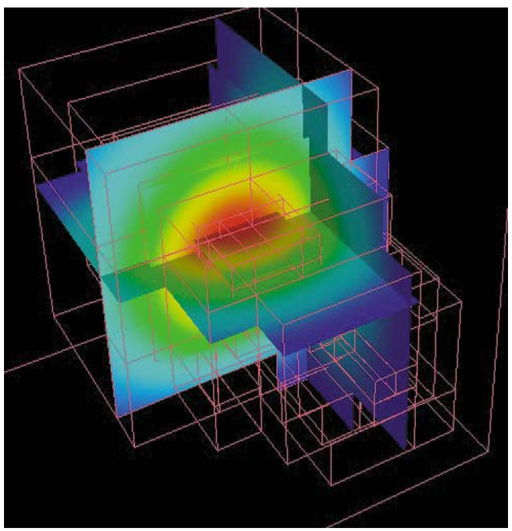
High Performance Computing Research Department

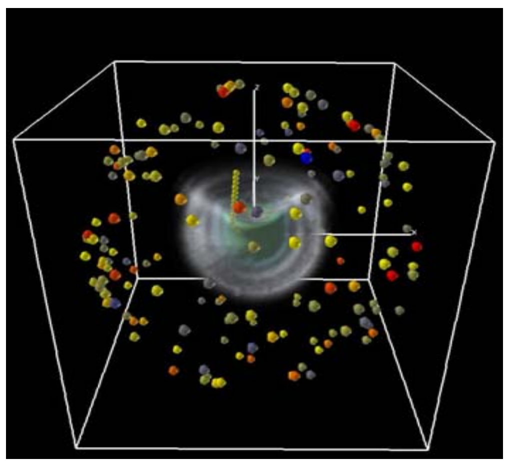




(a)







Juan Meza, Department Head

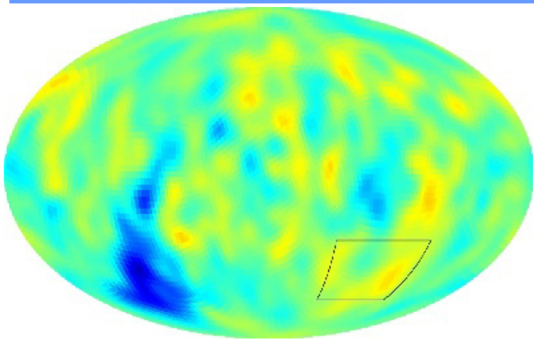
Groups:

- Applied Numerical Algorithms
- Center for Computational Sciences and Engineering
- Future Technologies
- Imaging and Informatics
- Mathematics
- Scientific Computing
- Scientific Data Management
- Visualization

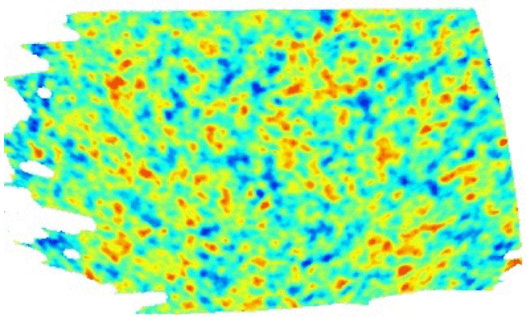
Total Staff: 122

The High Performance Computing Research Department conducts research and development in mathematical modeling, algorithm design, software implementation, and system architectures, and evaluates new and promising computer technologies.

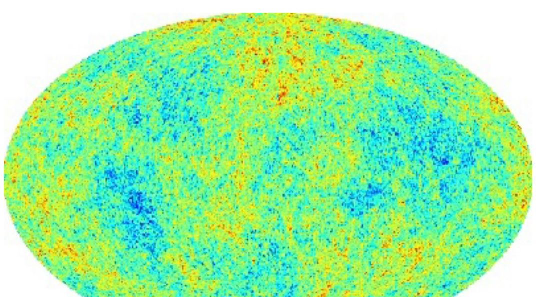
Cosmic Microwave Background Data Analysis - Taking The First Pictures



COBE All Sky, 10° Beam
(5000 x 3° Pixels)



BOOMERanG LDB 3% Sky, 10′ Beam
(100,000 x 7.5′ Pixels)

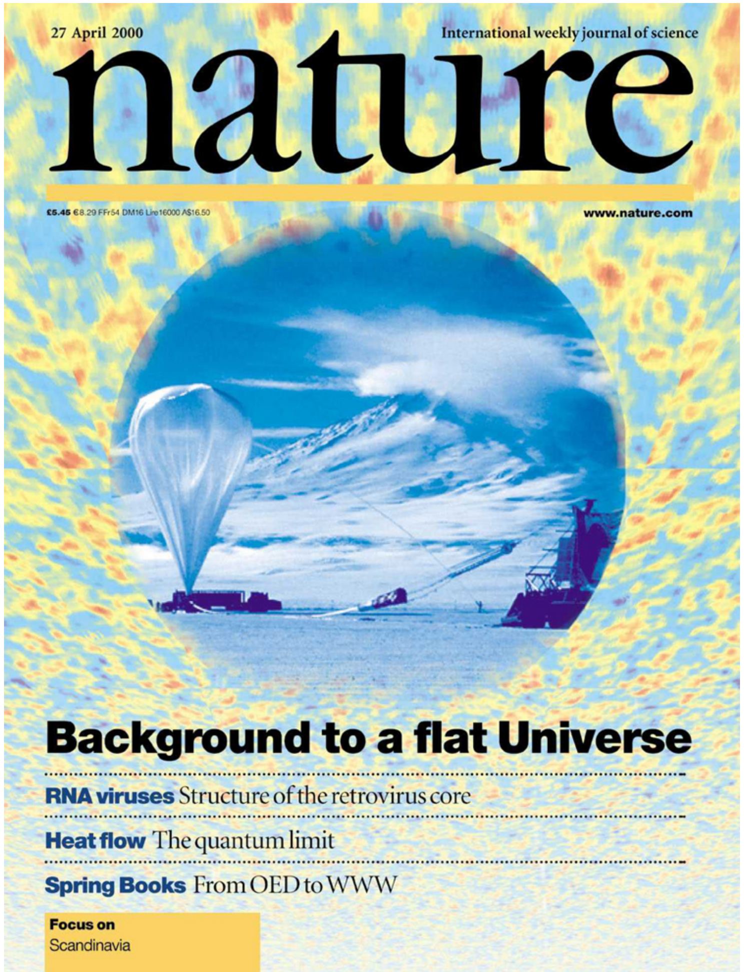


Simulated PLANCK All Sky, 5′ Beam
(12,000,000 x 3.75′ Pixels)

- The CMB is a snapshot of the Universe at 300,000 years & 3000 Kelvin.
- Primordial perturbations appear as tiny fluctuations in the CMB temperature and polarization.
- After 14 billion years of expansion, the CMB has cooled to 3 Kelvin with microKelvin anisotropies.
- CMB data analysis tries to distinguish these faint signals from the overwhelming noise, first in a map and then its angular power spectrum.

Julian Borrill, Scientific Computing, LBNL

Cosmic Microwave Background Data Analysis - Taking The First Pictures

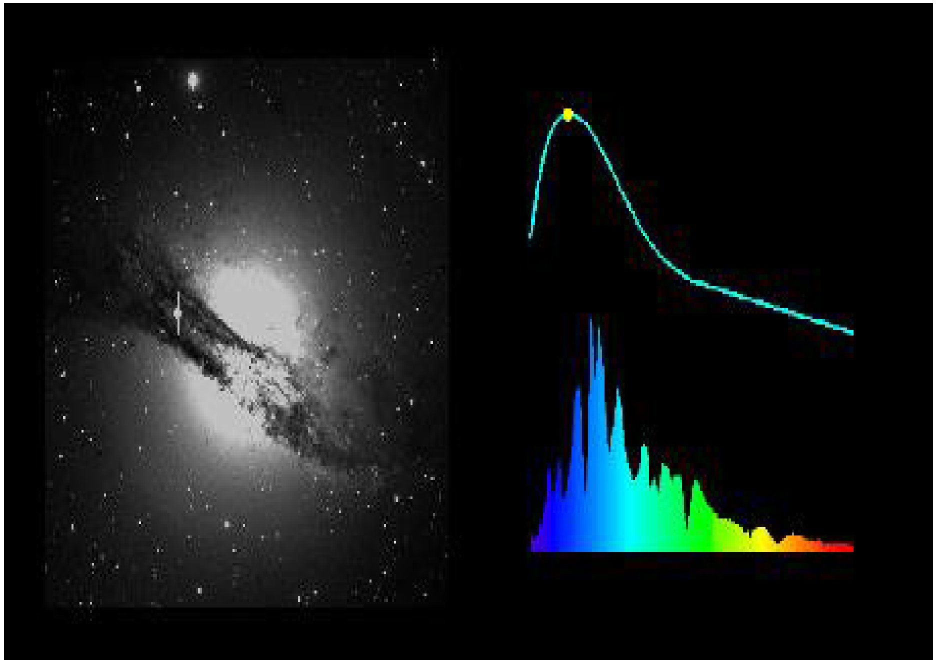


- As a problem in dense linear algebra, very efficient level 3 BLAS on massively parallel platforms allow the analysis of data such as the BOOMERanG Antarctic balloon observations.
- Such observations confirm that the Universe is flat, consisting of 65% dark energy, 30% dark matter and 5% ordinary matter, and constrain an entire class of fundamental quantum field theories.
- This exact analysis scales as N^3 in the number of sky pixels.
- Approximate Monte Carlo/FFT methods under development still require $O(10^{18})$ flops for a PLANCK size dataset, with excellent scalability but much lower efficiency

Julian Borrill, Scientific Computing, LBNL

Supernova Spectral Synthesis - Filling In The Light We Can't See

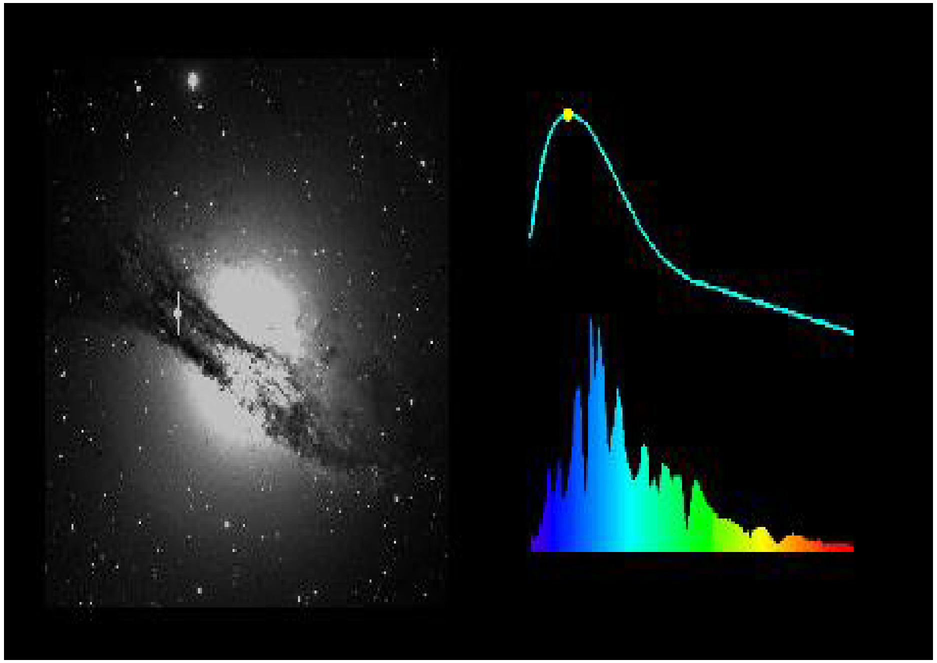
- The observed brightness and redshift of type Ia supernova determine the dynamical history of space - the expansion is accelerating!
- To understand a supernova observation we need its redshift, obtained by identifying features in its spectrum.



- Spectral observations have a limited frequency range; the greater the redshift of the supernova the higher the emission frequency of the observed spectrum.
- What does a supernova spectrum look like in the ultra-violet and beyond - only spectral synthesis simulations can tell us.

Peter Nugent, Scientific Computing, LBNL

Supernova Spectral Synthesis - Filling In The Light We Can't See



- Full 3D simulations are really required, scaling N^3 in the atmospheric resolution.
- Current 3D simulations use up to $O(10^2)$ gridpoints on a side, with grossly simplified physics.
- Physically realistic 3D simulations require $O(10^3 - 10^4)$ gridpoints on a side, and just begin to be possible at 100 Tflops.

Peter Nugent, Scientific Computing, LBNL

QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

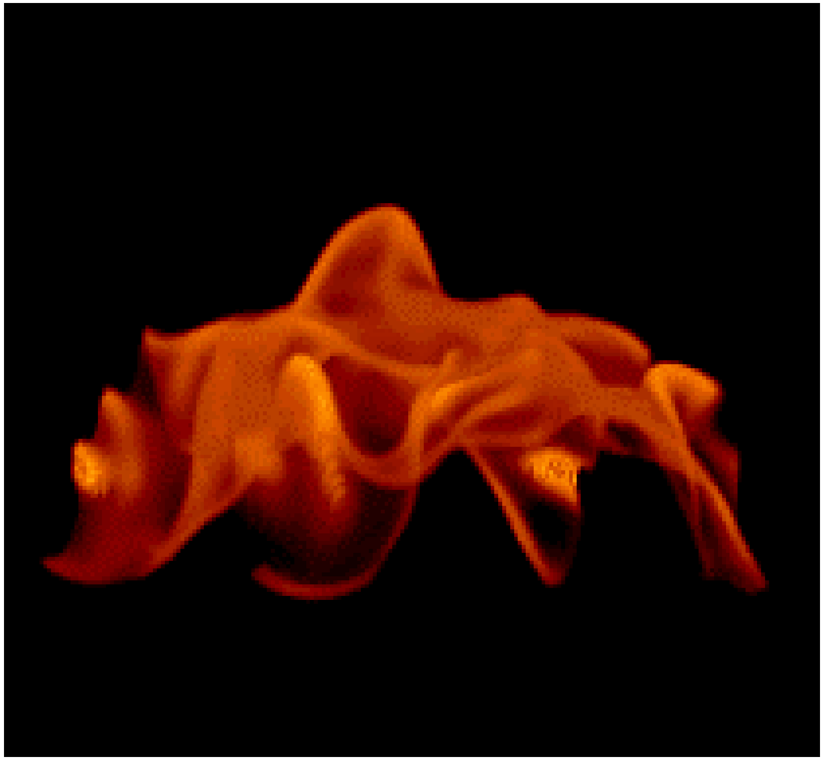
Climate modeling and predicting hurricane patterns

QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

- Tropical cyclones are not generally seen in integrations of global atmospheric general circulation models at climate model resolutions T42 (~300km)
- In CCM3 at T239 (50km), the lowest pressure attained is 995mb. No realistic cyclones are simulated.
- In high resolution simulations of the finite volume dynamics version of CAM2, strong tropical cyclones are common.

Michael Wehner, Scientific Computing, LBNL

Turbulent Methane Flame Sheet



Premixed methane flame sheet encountering isotropic turbulence, 19 species, 84 reactions, 8x8x16 mm, 256x256x512 grid points, Bell,Day, Grcar, Proc. Combust. Inst. 29, 2002, in press

- First fully resolved simulations of methane combustion with comprehensive chemistry
- Verified that flame acceleration is primarily but not entirely correlated with increased area of the flame sheet due to wrinkling
- Noted disparities in probability density function of species with respect to curvature and temperature
- Identified two mechanisms by which species concentrations are enhanced/depleted by wrinkling of the flame sheet

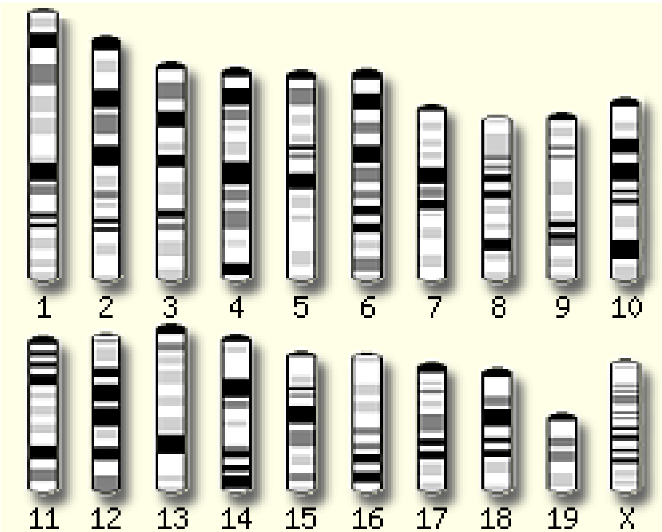
Turbulent Premixed V-Flame

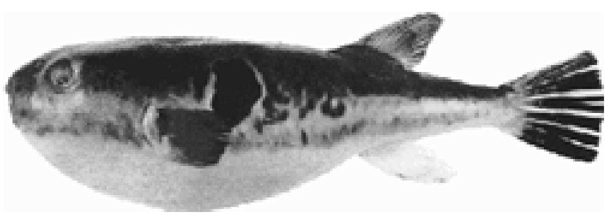
Experimental Turbulent V-Flame

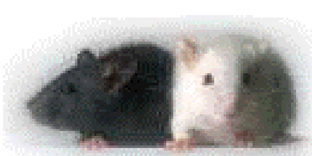
(photo courtesy R. K. Cheng, LBNL)

Recent Calculations

Assemby of Fugu genome



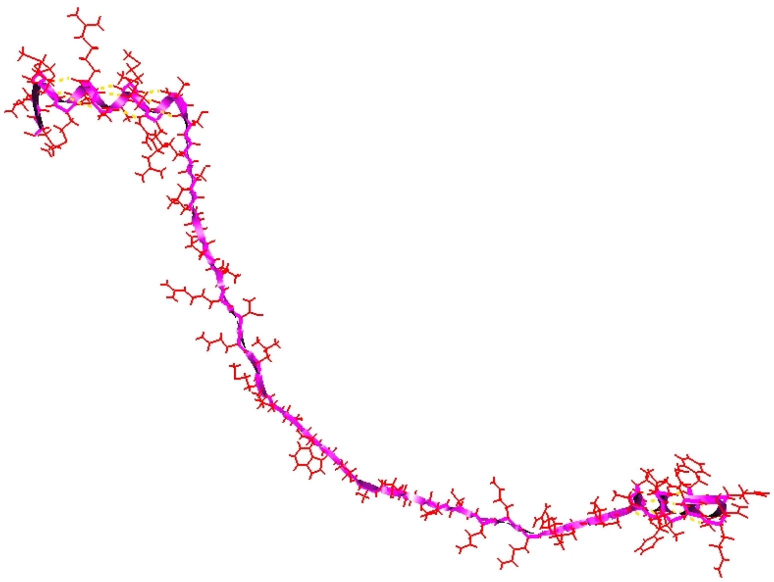


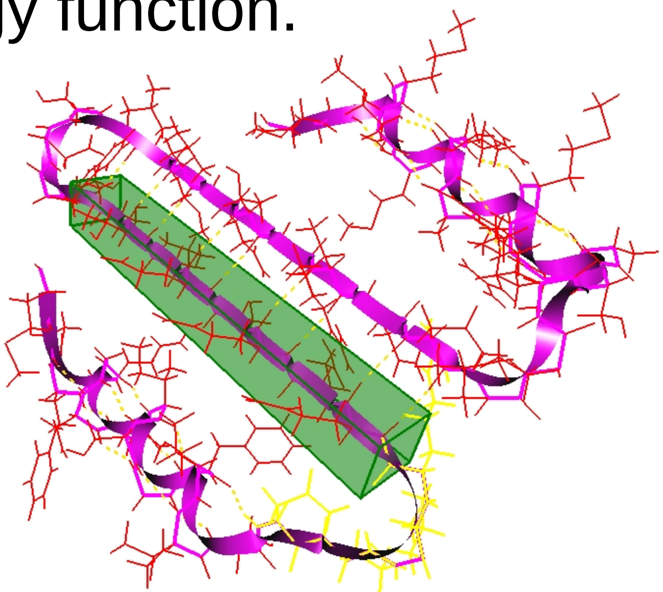


- Assembly of Fugu genome from 3.1 million reads, and initial preparation of mouse genome data.
- 75% of human genes have counterparts in Fugu genome
- Easier to find genes in Fugu because it has fewer noncoding (junk) DNA
- Led to prediction of 961 previously unidentified human genes
- Need new discrete math algorithms to study these problems

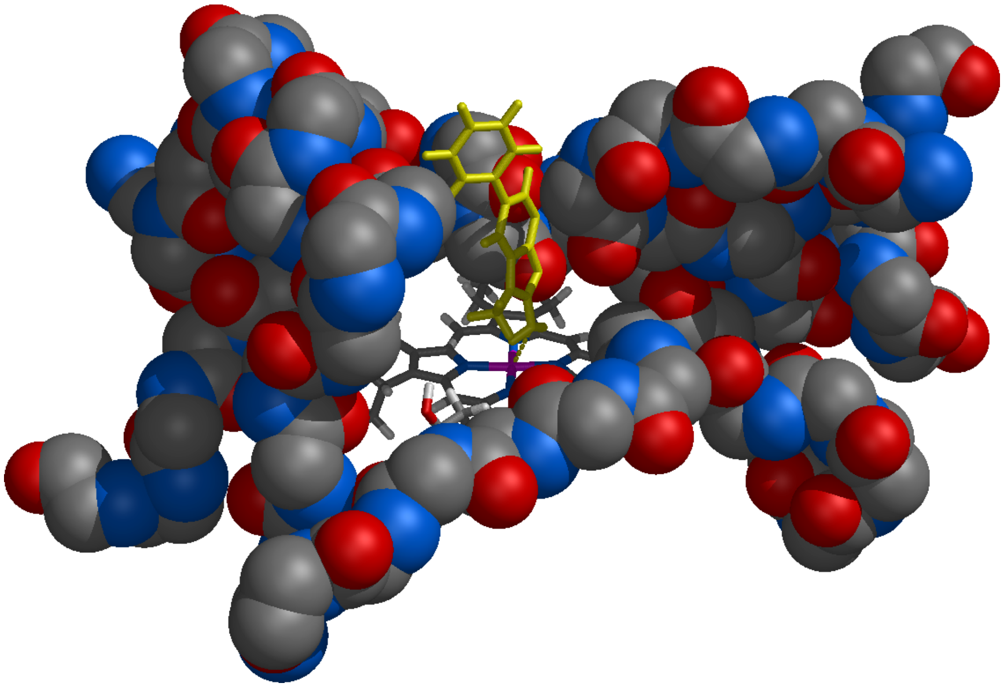
Accelerating Protein Structure Prediction

- Creating secondary structures: obtain predictions of α-helices and β-sheets from servers.
- Allow for interactive manipulation of one or more secondary structures using an inverse-kinematics constraint-based system.
- Added visualization of energy function.





Computational chemistry is used to design and study new molecules and drugs



Docking model for environmental carcinogen bound in *Pseudomonas Putida* cytochrome P450

- Drugs are typically small molecules which bind to and inhibit a target receptor
- Pharmaceutical design involves screening thousands of potential drugs
- A single new drug may cost over $500 million to develop
- The design process is time consuming (typically about 13 years)

The goal is to find the lowest energy for the molecule

$$\min \; E(r) = E_{Bonds} + E_{Angles} + E_{Dihedrals} + E_{NonBonded},$$

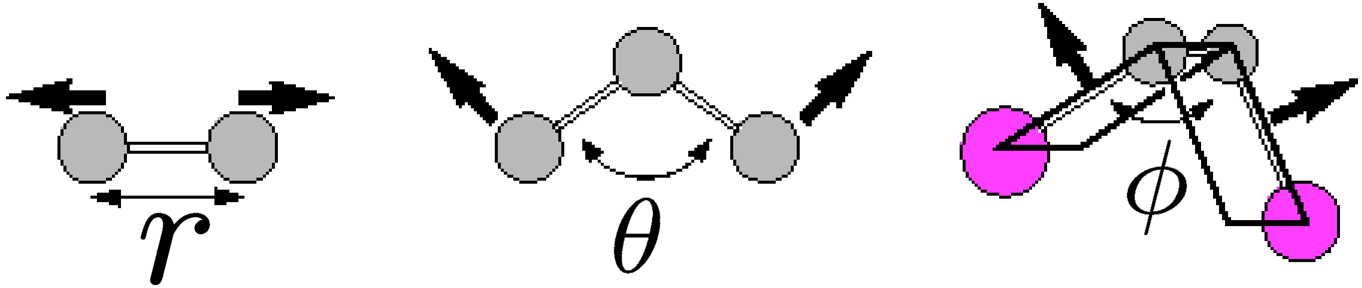
r denotes the distance between atoms,

E_{Bonds} is the energy between 2 atoms,

E_{Angles} is the energy between 3 atoms,

$E_{Dihedrals}$ is the energy between 4 atoms,

$E_{NonBonded}$ is the long distance energy



Source: http://www.chemistry.gatech.edu/faculty/collard/CHEM2311/hw/MM.pdf

Amber Function

$$E_{AMBER} = E_{Bonds} + E_{Angles} + E_{Dihedrals} + E_{NonBonded}$$

$$E_{Bonds} = \sum_{\text{Bonds}} K_{r_i} (r_i - \bar{r}_i)^2$$

$$E_{Angles} = \sum_{\text{Angles}} K_{\theta_i} \left(\theta_i - \bar{\theta}_i\right)^2$$

$$E_{Dihedrals} = \sum_{\text{Dihedrals}} K_{\phi_i} \left(1 + \cos(n_i \phi_i - \delta_i)\right)$$

$$E_{NonBonded} = \sum_i \sum_{i<j} \left(\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right)$$

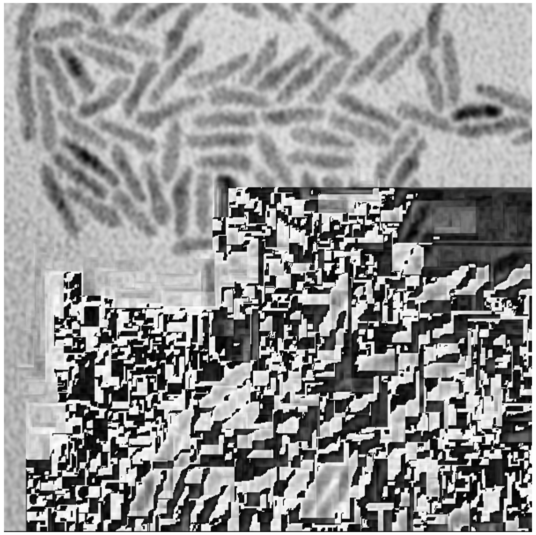
A Physical Approach to Protein Structure Prediction, Crivelli, et.al. Biophysical Journal, Vol 82, 2002.

Protein T162 from CASP5

QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

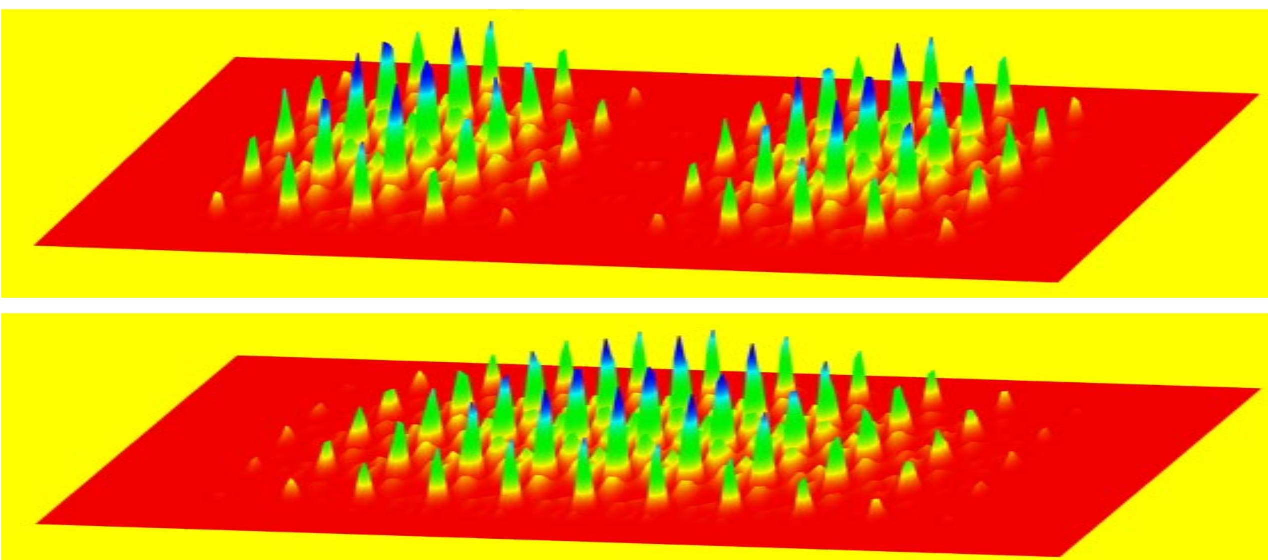
- Initial configuration created using ProteinShop (S. Crivelli)
- Energy minimization computed using OPT++/LBFGS
- Final average RMSD change was 3.9 Å
- Total simulation took approximately 32 hours

Computational Nanoscience



CdSe quantum rods

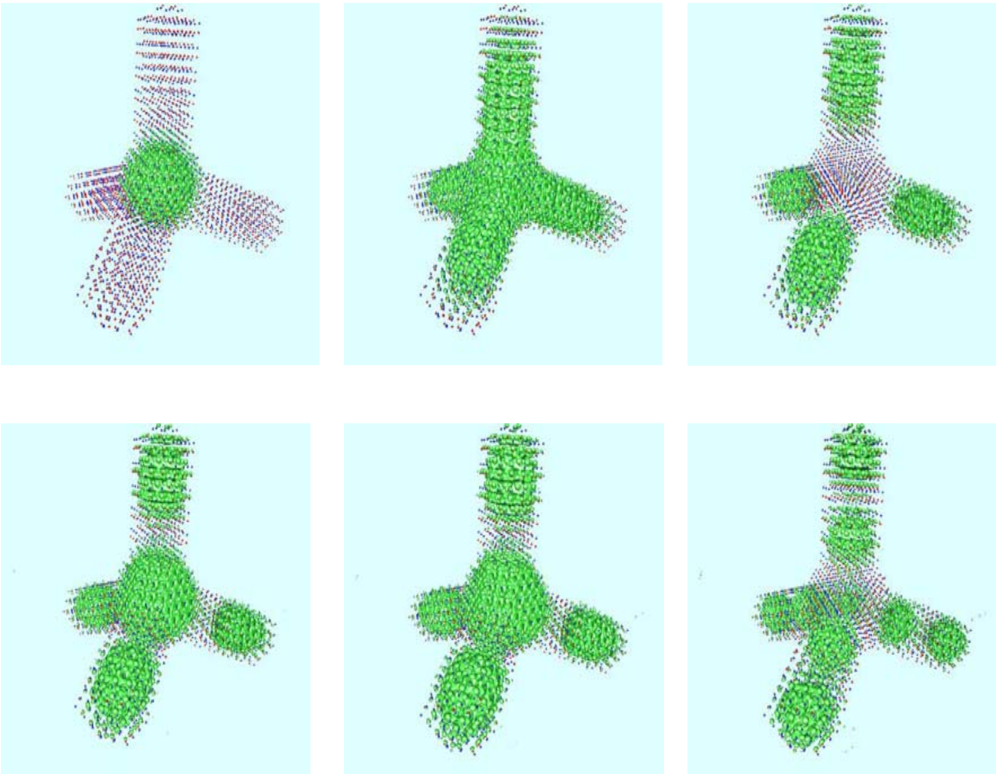
- The electronic structure and optical properties change with the shape of the quantum rods
- The thousand atom quantum rods can be calculated using the planewave pseudopotential method
- Software exists to simulate nanosystems and compare with experimental electronic and optical results



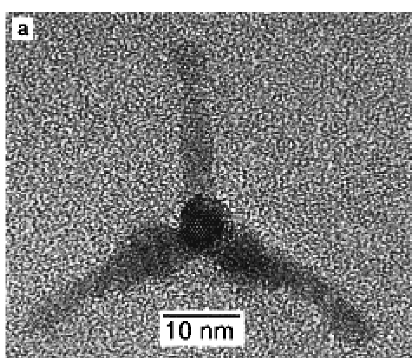
The electron wavefunctions of quantum rods

Andrew Canning, Lin-Wang Wang, Scientific Computing, LBNL

CdSe tetrapod conduction band wavefunctions

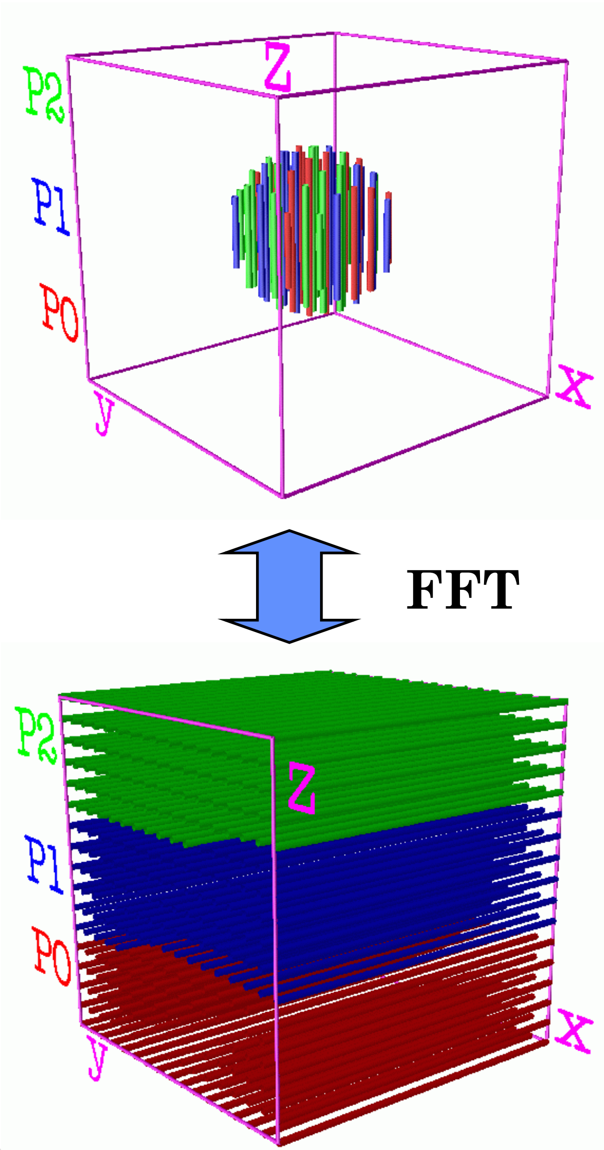


- First 6 electron wavefunctions of a CdSe tetrapod quantum structure
- System contains ~1000 atoms
- Solution of large eigenvalue problem is required

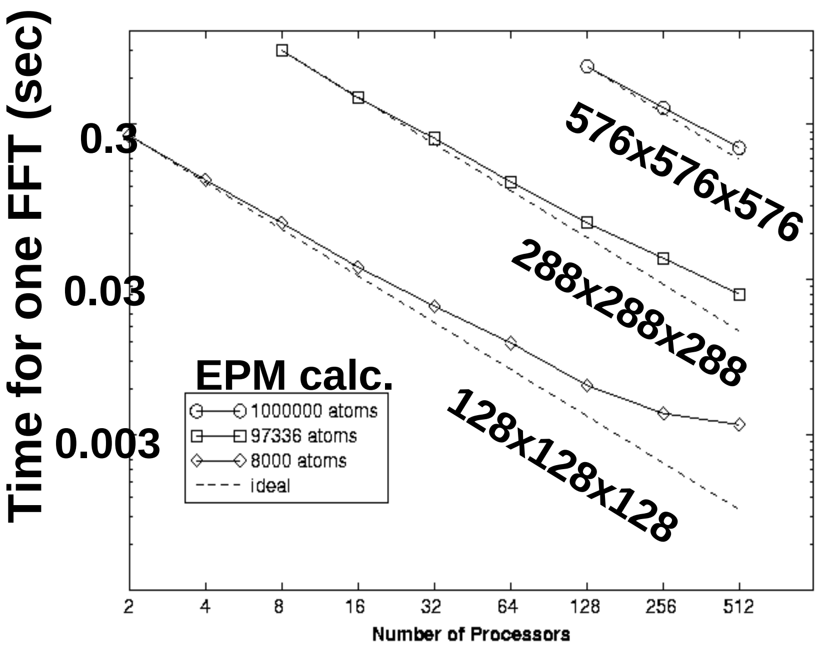


Andrew Canning, Lin-Wang Wang,
Scientific Computing, LBNL

PARATEC Kernel: A parallel Fast Fourier Transformation code



- Specially designed for Plane Wave electronic structure calculations
- Work load balance
- Memory balance
- Minimum communication



Summary

- Computational sciences and mathematics are increasingly being used to explore science problems
  - Climate, Combustion: PDEs, ODEs, …
  - Chemistry, Material Sciences: FFTs,
  - Nanosciences: Linear Algebra/Eigenvalues
  - Biology: Discrete math and combinatorics
  - Everywhere: Nonlinear equations and optimization

- Mathematical hurdles must be overcome to solve real world problems
- The (correct) use of mathematics can accelerate the discovery of science through new capabilities

Thank you

Q&A